

Bayesian inference in a parametric  
counting process model - an asymptotic result

by

Terje Aven

Abstract

We are in this paper concerned with Bayesian inference in a counting process model where the intensities depend on an unknown parameter. In particular, the model gives a unified approach to Bayesian inference for a large number of parametric failure time models with censoring. Asymptotic properties of the Bayes estimator (under quadratic loss function) in the "exponential model" are studied.

Key words: counting processes, intensities, martingales, Bayesian inference, survival analysis, life testing, censoring, exponential life distributions, asymptotic theory.

## 1. Introduction

Let  $\{N_1(t), \dots, N_k(t), t \in [0, \infty)\}$  be a multivariate counting process adapted to a history  $\{F_t\}$  and let  $\{P_\theta, \theta \in \Theta\}$  be a family of probabilities. Assume that  $N_i(t) < \infty, 0 < t < \infty$ , and that

$$(1.1) \quad \left\{N_i(t) - \int_0^t \lambda_i(\theta, s) ds\right\} \text{ is a } (P_\theta, F_t)\text{-local martingale,}$$

i.e.  $\{N_i(t)\}$  has  $(P_\theta, F_t)$ -intensity  $\{\lambda_i(\theta, t)\}$ .

We remark that under certain conditions (see Aalen (1978) p.705, Brémaud (1981) p.28 and Aven (1983)) (1.1) is equivalent to

$$\lim_{h_n \downarrow 0} \frac{1}{h_n} E_\theta[N_i(t+h_n) - N_i(t) | F_t] = \lambda_i(\theta, t^+).$$

We assume that

$$F_t = F_0 \vee \sigma(N_i(u), u \leq t, i=1, 2, \dots, k)$$

and that

$$(1.2) \quad \lambda_i(\theta, t) = \mu_i(\theta, t) Y_i(t)$$

(the assumption (1.2) is no restriction).

Furthermore, we assume that there exists a probability  $Q$  such that

$$(1.3) \quad P_\theta(A) = \int_A L_0(\theta) dQ, \quad A \in F_0$$

( $L_0(\theta)$  is  $F_0$ -measurable) and such that

$$\{N_i(t)\} \text{ has } (Q, F_t)\text{-intensity } \{Y_i(t)\}.$$

The statement (1.3) says that the restriction of  $P_\theta$  to  $F_0$  is absolutely continuous with respect to the restriction of  $Q$  to  $F_0$  and the Radon-Nikodym derivative equals  $L_0(\theta)$ . In applications we usually have  $L_0(\theta) = 1$ .

The true value  $\theta_0$  of  $\theta$  is supposed to be unknown. Our initial uncertainty about  $\theta_0$  is expressed by a prior distribution  $\mu$  on  $\Theta$ . We seek the posterior distribution given  $F_T$ , where  $T$  is a finite  $F_t$ -stopping time.

In Section 2 we present the model. By applying general counting process results we show that

$$(1.4) \quad P_\theta(A) = \int_A L(\theta, T) dQ, \quad A \in F_T,$$

where the likelihood  $L$  is given by

$$\begin{aligned} L(\theta, T) &= L_0(\theta) \left( \prod_{i=1}^k \prod_{j=1}^{N_i(T)} \mu_i(\theta, D_{ij}) \right) \exp \left\{ \sum_{i=1}^k \int_0^T (1 - \mu_i(\theta, s)) Y_i(s) ds \right\} \\ &= L_0(\theta) \exp \left\{ \sum_{i=1}^k \int_0^T \ln \mu_i(\theta, s) dN_i(s) + \sum_{i=1}^k \int_0^T (1 - \mu_i(\theta, s)) Y_i(s) ds \right\}; \end{aligned}$$

here  $D_{ij}$  is the  $j$ -th jump time of  $N_i$ .

The statement (1.4) says that the restriction of  $P_\theta$  to  $F_T$  is absolutely continuous with respect to the restriction of  $Q$  to  $F_T$  and the Radon-Nikodym derivative equals  $L(\theta, T)$ .

In Section 3 we formulate the Bayes set-up. We are here more or less copying known mathematical statistical theory. We show that the posterior distribution given  $F_T$  equals

$$\frac{\int_B L(\theta, T) d\mu(\theta)}{\int_\Theta L(\theta, T) d\mu(\theta)}.$$

The above model is quite general, in particular, it includes a large number of failure time models with censoring. As an illustration we shall give a special case from life testing. At time  $t = 0$   $n_i$  units are put on test in environment  $i$ ,  $i = 1, 2, \dots, k$ . The lifelengths (time to failures) of the units in environment  $i$  are random variables with failure rate  $\alpha_i(\theta, t)$ ; all lifelengths are independent. The testing in environment  $i$  is stopped at a finite stopping time  $T_i$ , based on the available information ( $T_i$  may for instance be a constant). Now, let  $N_i(t)$  denote the number of observed failures in environment  $i$  in  $[0, t]$ . Then  $\{N_1(t), \dots, N_k(t)\}$  is a multivariate counting process and it can be shown that  $\{N_i(t)\}$  has  $F_t = \sigma(N_i(u), u \leq t, i = 1, \dots, k)$ -intensity  $\{\alpha_i(\theta, t) Y_i(t)\}$ , where  $Y_i(t)$  is the number of units on test (at risk) in environ-

ment  $i$  just before time  $t$ . The true value  $\theta_0$  of  $\theta$  is supposed to be unknown and our initial uncertainty about  $\theta_0$  is expressed by a distribution  $\mu$ . We seek the posterior distribution given the data, i.e., the posterior distribution given  $F_T$ , where  $T = \max T_i$ .

The intensity form  $\alpha_i(t)Y_i(t)$ , where  $\alpha_i$  is a deterministic function and  $Y_i$  is an  $F_t$ -adapted process, is Aalen's (1978) well known intensity form. This form is very often in force when failure time models with censoring is formulated in a counting process framework, see e.g. Aalen (1975,1978), Gill (1980) and Borgan (1983). It should be noted that if the units are replaced by new ones at failures in the above example, then Aalen's intensity form does not hold. However, the example is still covered by our model.

Our model has much in common with Rebolledo's (1978) model (M1a,...M4a) (when ignoring nullsets our model is a special case of Rebolledo's (1978) general statistical model (M1,...,M4)). As opposed to Rebolledo (1978) we avoid the assumption " $E_Q L(\theta, T) = 1$ ".

In Section 4 we study asymptotic properties of the Bayes estimator (under quadratic loss function)

$$\hat{\theta}^n = \frac{\int_{\theta} \theta L^n(\theta, T^n) p(\theta) d\theta}{\int_{\theta} L^n(\theta, T^n) p(\theta) d\theta}$$

of  $\theta_0$  with respect to the probabilities  $P_{\theta_0}^n$ , assuming that we have a sequence of models, indexed by  $n = 1, 2, \dots$ , with  $k = 1$ ,  $\theta = (0, \infty)$ ,  $L_0^n(\theta)$  not dependent on  $\theta$ ,  $\lambda_1^n(\theta, t) = g(\theta)Y_1^n(t)$ , where  $g(\theta) = \theta$  (for all  $\theta$ ) or  $g(\theta) = \theta^{-1}$  (for all  $\theta$ ), and  $d\mu(\theta) = p(\theta)d\theta$ .

In life testing applications the intensity form  $g(\theta)Y_1^n(t)$  corresponds to exponential distributions with failure rate  $g(\theta)$  ( $=\theta$  or  $\theta^{-1}$ ) and mean life to failure  $g(\theta)^{-1}$  ( $=\theta^{-1}$  or  $\theta$ ) for the lifelengths of the units. The index  $n$  usually represents the total number of units which are put on test at  $t = 0$ .

Under certain conditions we show that  $\sqrt{n}(\hat{\theta}^n - \theta_0)$  is asymptotically normal with mean 0 and a variance  $\beta^{-1}$  (which does not depend on the prior density  $p$ ). This implies, in particular, that  $\hat{\theta}^n$  is consistent. The estimator  $\hat{\theta}^n$  takes a very simple form if  $g(\theta) = \theta$  and  $p$  is a gamma density or if  $g(\theta) = \theta^{-1}$  and  $p$  is an inverted gamma density (see Section 4). In fact the gamma (inverted gamma) prior density is a natural conjugate prior density (cf. Barlow and Proschan (1979, 1980)).

In Section 5 a special case of the set-up of Section 4 is considered.

Asymptotic properties of "nonparametric estimators" and maximum likelihood estimators in counting process models have been studied by for example Aalen (1975, 1978), Rebolledo (1978) and Borgan (1983).

We will use some basic definitions from the theory of stochastic processes without further comment. Our references as regards this theory are Brémaud and Jacod (1977), Gill (1980) and Brémaud (1981).

## 2. The counting process model

Let  $(\Omega, \Sigma, Q)$  be a probability space and let  $F_0$  be a sub- $\sigma$ -field of  $\Sigma$ . Let  $\{N_1(t), N_2(t), \dots, N_k(t), t \in [0, \infty)\}$  be a  $k$ -variate counting process on  $(\Omega, \Sigma)$  with values in  $\{0, 1, 2, \dots\}^k$  and define

$$F_t = F_0 \vee \sigma(N_i(s), s \leq t), \quad 0 \leq t < \infty.$$

We remark that  $\{F_t\}$  is right-continuous (see Corollary A.2.1, Appendix 2 of Gill (1980)). Furthermore, let for  $i = 1, 2, \dots, k$   $\{Y_i(t), t \in [0, \infty)\}$  be a non-negative  $F_t$ -progressively measurable process such that

$$\int_0^t Y_i(s) ds < \infty, \quad 0 \leq t < \infty.$$

Assume that for  $i = 1, 2, \dots, k$

$\{N_i(t)\}$  has  $(Q, F_t)$ -intensity  $\{Y_i(t)\}$ ,

i.e.

$$E_Q \int_0^\infty X(t) dN_i(t) = E_Q \int_0^\infty X(t) Y_i(t) dt$$

for all non-negative  $F_t$ -predictable processes  $\{X(t)\}$ , or equivalently

$\{N_i(t) - \int_0^t Y_i(s) ds\}$  is a  $(Q, F_t)$ -local martingale.

Let  $(\theta, B)$  be a measurable space and let for  $i = 1, 2, \dots, k$   $\mu_i(\theta, t, \omega)$  be a non-negative function on  $\theta \times [0, \infty) \times \Omega$  such that

$$(2.1) \quad \mu_i \text{ is } B \times P(F_t)\text{-measurable,}$$

here  $P(F_t)$  is the  $F_t$ -predictable  $\sigma$ -field over  $[0, \infty)$  ( $P(F_t)$  is generated by all processes which are adapted and left-continuous).

It should be noted that (2.1) implies that for each  $\theta$

$$(2.2) \quad \{\mu_i(\theta, t)\} \text{ is an } F_t\text{-predictable process.}$$

Assume that for  $i = 1, 2, \dots, k$  and  $\theta \in \theta$

$$\int_0^t \mu_i(\theta, s) Y_i(s) ds < \infty, \quad 0 < t < \infty.$$

Define

$$\lambda_i(\theta, t) = \mu_i(\theta, t) Y_i(t), \quad i = 1, 2, \dots, k, \quad \theta \in \theta, \quad 0 < t < \infty.$$

Let  $P_\theta$ ,  $\theta \in \theta$ , be probabilities on  $(\Omega, \Sigma)$ . Assume that

$$(2.3) \quad P_\theta(A) = \int_A L_0(\theta) dQ, \quad A \in F_0,$$

where

$$(2.4) \quad L_0(\theta) \text{ is a non-negative } B \times F_0\text{-measurable function on } \theta \times \Omega.$$

Furthermore, assume that

$$(2.5) \quad \{N_i(t)\} \text{ has } (P_\theta, F_t)\text{-intensity } \{\lambda_i(\theta, t)\}, \quad i = 1, 2, \dots, k.$$

Define for each  $\theta \in \Theta$  the process  $\{L(\theta, t)\}$ ,  $t \in [0, \infty)$  by

$$L(\theta, t) = L_0(\theta) \exp \left\{ \sum_{i=1}^k \int_0^t \left[ \ln \mu_i(\theta, s) dN_i(s) + (1 - \mu_i(\theta, s)) Y_i(s) ds \right] \right\}.$$

By using (2.1) and (2.4) it can be shown that  $L(\theta, T)$  is a  $B \times F_T$ -measurable function on  $\Theta \times \Omega$  for each  $T \in \mathcal{T}$ , where  $\mathcal{T}$  is the collection of all finite  $F_t$ -stopping times, see the appendix.

We end this section by formulating and proving a proposition which says that the statistical space (experiment)  $(\Omega, F_T, P_\theta, \theta \in \Theta)$  is dominated by  $Q$  (it is understood that the probabilities are restricted to  $(\Omega, F_T)$ ) and that the likelihood is  $L(\theta, T)$ .

Proposition 2.1

$$(2.6) \quad P_\theta(A) = \int_A L(\theta, T) dQ, \quad \theta \in \Theta, \quad A \in F_T, \quad T \in \mathcal{T}.$$

Remark. The statement (2.6) also holds if (2.1) is replaced by (2.2) and (2.4) is replaced by " $L_0(\theta)$  is an  $F_0$ -measurable function for each  $\theta$ " (this is seen from the proof below).

Proof. Let  $T \in \mathcal{T}$ . It is not difficult to see that

$$\{N_i(t \wedge T)\} \text{ has } (Q, F_{t \wedge T})\text{-intensity } \{Y_i(t) I(t < T)\}$$

and

$$\{N_i(t \wedge T)\} \text{ has } (P_\theta, F_{t \wedge T})\text{-intensity } \{\lambda_i(\theta, t) I(t < T)\}.$$

Note that  $F_{t \wedge T} = F_0 \vee \sigma(N_i(s \wedge T), s < t, i = 1, 2, \dots, k)$  (see Gill (1980) Appendix 2, Corollary A.2.1). Now, applying the results stated in Brémand and Jacod (1977) pp. 388-389 (let  $N_t^x = N_x(t \wedge T)$ ,  $F_t = F_{t \wedge T}$ ,  $Y(t, x) = \mu_x(\theta, t)$  and  $\tilde{A}([0, t] \times \{x\}) = \int_0^t Y_x(s) I(s < T) ds$ )

we find that the following statements hold:

$$(2.7) \quad \{L(\theta, t \wedge T), t \in [0, \infty)\} \text{ is a uniformly integrable } (Q, F_{t \wedge T})\text{-martingale, } \theta \in \Theta,$$

and

$\{N_i(t \wedge T)\}$  has  $(P_\theta^*, F_{t \wedge T})$ -intensity  $\{\lambda_i(\theta, t)I(t \leq T)\}$ ,

where  $P_\theta^*$  is defined by

$$P_\theta^*(A) = \int_A L(\theta, T) dQ, \quad A \in \Sigma,$$

$i = 1, 2, \dots, k$ , and  $\theta \in \Theta$ .

The statement (2.6) follows if we can prove that  $P_\theta = P_\theta^*$  on  $F_T$ . Let  $A \in F_0$ . Using (2.7) and (2.3) we obtain

$$\begin{aligned} P_\theta^*(A) &= \int_A L(\theta, T) dQ = \int_A E_Q[L(\theta, T) | F_0] dQ \\ &= \int_A L(\theta, 0) dQ = \int_A L_0(\theta) dQ = P_\theta(A). \end{aligned}$$

Hence  $P_\theta = P_\theta^*$  on  $F_0$ . From the uniqueness theorem stated in Brémaud and Jacod (1977) p. 388 we can conclude that  $P_\theta = P_\theta^*$  on  $F_{\infty \wedge T} = F_T$ . This completes the proof of the proposition.

### 3. The prior and the posterior distributions

In this section we formulate the Bayes set-up. We are here more or less copying known mathematical statistical theory, cf. e.g. Barra (1981) pp. 7-8. Consider the model described in Section 2. The true value  $\theta_0$  of  $\theta$  is supposed to be unknown. Our initial uncertainty about  $\theta_0$  is expressed by a prior distribution  $\mu$  on  $\Theta$ . We seek the posterior distribution on  $\theta$  given  $F_T$ ,  $T \in \mathcal{T}$ .

Define a probability  $\tilde{P}$  on  $(\Omega, F_\infty)$  by

$$(3.1) \quad \tilde{P}(A) = \int_\Theta P_\theta(A) d\mu(\theta).$$

The probability  $\tilde{P}$  is our subjective probability on  $F_\infty$  obtained by weighting  $P_\theta(\cdot)$  by  $\mu$ . If  $\tilde{\theta}$  is a random variable with distribution  $\mu$  and  $P_\theta(A)$  is considered as the conditional probability of the event  $A$  given  $\tilde{\theta} = \theta$ , then  $\tilde{P}(A)$  is simply the unconditional probability. Let for  $B \in \mathcal{B}$  and  $T \in \mathcal{T}$



$$(3.2) \quad \pi_T(B) = \frac{\int_B L(\theta, T) d\mu(\theta)}{\int_{\Theta} L(\theta, T) d\mu(\theta)},$$

where  $0/0 \stackrel{\text{def}}{=} 1$  and  $a/\infty \stackrel{\text{def}}{=} 1$ . We see that  $\pi_T(B, \omega)$  is a probability on  $(\Theta, \mathcal{B})$  for each  $\omega$ , and an  $F_T$ -measurable function for each  $B$  (remember that  $L$  is a  $\mathcal{B} \times F_T$ -measurable function on  $\Theta \times \Omega$ ). We shall show that  $\pi_T$  is a posterior distribution given  $F_T$ , i.e. for all  $A \in F_T$  and all  $B \in \mathcal{B}$  we have ((3.3) is motivated below)

$$(3.3) \quad \int_A \pi_T(B) d\tilde{P} = \int_B P_{\theta}(A) d\mu(\theta),$$

cf. Barra (1981) p. 8. Now, by combining (3.1) and (2.6) and by changing the order of integration we find that

$$(3.4) \quad \tilde{P}(A) = \int_A \int_{\Theta} L(\theta, T) d\mu(\theta) dQ, \quad A \in F_T.$$

Using (3.4) and (2.6) we obtain

$$\begin{aligned} \int_A \pi_T(B) d\tilde{P} &= \int_A \pi_T(B) \int_{\Theta} L(\theta, T) d\mu(\theta) dQ = \int_A \int_B L(\theta, T) d\mu(\theta) dQ \\ &= \int_B \int_A L(\theta, T) dQ d\mu(\theta) = \int_B P_{\theta}(A) d\mu(\theta), \end{aligned}$$

which proves (3.3).

We formulate this result as a theorem.

**Theorem 3.1.** The posterior distribution on  $\theta$  given  $F_T$  is given by (3.2),  $T \in \mathcal{T}$ .

To motivate (3.3), suppose that  $\tilde{\theta}$  is a random variable with distribution  $\mu$  on  $\Theta$ , and consider  $P_{\theta}(A)$  as the conditional probability of  $A$  given  $\tilde{\theta} = \theta$ . Then we have for  $A \in F_T$  and  $B \in \mathcal{B}$

$$\int_A \tilde{P}[\tilde{\theta} \in B | F_T] d\tilde{P} = \tilde{P}(A \cap \{\tilde{\theta} \in B\}) = \int_B P_{\theta}(A) d\mu(\theta).$$

It follows that

$$\pi_T(B) = \tilde{P}[\tilde{\theta} \in B | F_T] \quad \tilde{P} - \text{a.s.},$$

which shows that  $\pi_T(B)$  in fact is a posterior distribution given  $F_T$ .

Finally in this section we shall give some results which are related to the set-up of Section 4. For the sake of convenience we shall drop a.s. phrases.

Assume that  $k = 1$ ,  $\theta = (0, \infty)$ ,  $\mathcal{B}$  = the class of Borel sets on  $(0, \infty)$ ,  $L_0(\theta)$  does not depend on  $\theta$ ,  $\mu$  is a gamma distribution with shape parameter  $c$  and scale parameter  $d$ , i.e.

$$\mu(B) = \int_B [d^c \theta^{c-1} e^{-d\theta} / \Gamma(c)] d\theta,$$

and

$$(3.5) \quad \mu(\theta, t) = \theta, \quad \text{i.e.} \quad \lambda(\theta, t) = \theta Y(t)$$

(we write  $\mu$ ,  $\lambda$  and  $Y$  in stead of  $\mu_1$ ,  $\lambda_1$  and  $Y_1$ , respectively). It is not difficult to see that the posterior distribution  $\pi_T$  is also a gamma distribution, the shape parameter and scale parameter changed to  $c + N(T)$  and  $d + \int_0^T Y(s)ds$ , respectively. Thus the gamma prior is a natural conjugate prior. The posterior mean,  $\int_0 \theta \pi_T(d\theta)$ , i.e. the Bayes estimator of  $\theta_0$  under quadratic loss function (an estimator  $\hat{\theta}$  is called a Bayes estimator under quadratic loss function if  $\hat{\theta}$  minimizes  $\int_0 (\theta - x)^2 \pi_T(d\theta)$ ) is given by

$$\int_0 \theta \pi_T(d\theta) = \frac{c+N(T)}{d+\int_0^T Y(s)ds}.$$

Now suppose (3.5) is replaced by

$$\mu(\theta, t) = \theta^{-1}, \quad \text{i.e.} \quad \lambda(\theta, t) = \theta^{-1} Y(t).$$

Then the inverted gamma distribution  $(\int_B [d^c \theta^{-(c+1)} e^{-d/\theta} / \Gamma(c)] d\theta)$  is a natural conjugate prior. In the posterior distribution the parameters are  $c + N(T)$  (in place of  $c$ ) and  $d + \int_0^T Y(s)ds$  (in place of  $d$ ). The mean takes the form

$$\frac{d+\int_0^T Y(s)ds}{c+N(T)-1}$$

(cf. Barlow and Proschan (1979, 1980)).

#### 4. An asymptotic result

Consider the model described in Section 2 with  $k = 1$ .

Assume:

A1.  $(\Omega, \Sigma, Q)$  is a complete probability space and  $F_0$  includes all  $Q$ -null sets.

A2. The set  $\theta$  equals the interval  $(0, \infty)$  and  $\mathcal{B}$  is the class of Borel sets on  $\theta$ .

A3.  $L_0(\theta)$  does not depend on  $\theta$ .

A4.  $\mu(\theta, t) = g(\theta)$  (i.e.  $\lambda(\theta, t) = g(\theta)Y(t)$ ),  
where  
 $g(\theta) = \theta$  or  $g(\theta) = \theta^{-1}$ .

A5. The prior distribution  $\mu$  on  $\theta$  has a density  $p$  with respect to Lebesgue measure which is continuous and positive at the true value  $\theta_0$  of  $\theta$ .

Furthermore,

$$\int_{\theta} \theta p(\theta) d\theta < \infty.$$

Note that  $Q = P_1$ .

Remark. Our analysis in this section need not be changed even though  $g$  does not take the form  $\theta$  or  $\theta^{-1}$ . We must, however, require that  $g$  is strictly non-decreasing or strictly non-increasing,  $g'$  exists, and  $g'$  is continuous at  $\theta_0$ .

Some of the assumptions made in Section 2 are not directly used in this section, for example the assumption that  $F_t$  takes the form  $F_0 \vee \sigma(N(s), s \leq t)$ .

Assume now that we have given a sequence of models of the above form, indexed by  $n = 1, 2, \dots$ , where the prior density  $p$  and the function  $g$  are the same in all the models, and where a

$T^n \in T^n$  is chosen. Assume that the following conditions are satisfied:

A6. There exists a measurable and deterministic function  $y(t)$  such that

$$(4.1) \quad 0 < b \stackrel{\text{def}}{=} \int_0^{\infty} y(s) ds < \infty,$$

$$(4.2) \quad \frac{1}{n} \int_0^{T^n} Y^n(s) ds \xrightarrow{P} b$$

and for all  $\varepsilon > 0$

$$(4.3) \quad \lim_{t \uparrow \infty} \limsup_{n \rightarrow \infty} P_{\theta_0}^n \left\{ \frac{1}{n} \int_{t \wedge T^n}^{T^n} Y^n(s) ds > \varepsilon \right\} = 0;$$

here  $\xrightarrow{P}$  means convergence in probability relative to the probabilities  $P_{\theta_0}^n$ .

Remark. In applications (see Section 5), (4.2) is often established by showing that

$$(4.4) \quad \frac{1}{n} \int_0^{t \wedge T^n} Y^n(s) ds \xrightarrow{P} \int_0^t y(s) ds, \quad 0 < t < \infty$$

(it seen by writing  $\frac{1}{n} \int_0^{T^n} Y^n(s) ds - \int_0^{\infty} y(s) ds = \left[ \frac{1}{n} \int_0^{t \wedge T^n} Y^n(s) ds - \int_0^t y(s) ds \right] + \frac{1}{n} \int_{t \wedge T^n}^{T^n} Y^n(s) ds - \int_t^{\infty} y(s) ds$  that (4.4) together with (4.1) and (4.3) imply (4.2)).

For the sake of convenience we shall in the following write  $P\{ \}$  instead of  $P_{\theta_0}^n\{ \}$  and drop all  $P_{\theta_0}^n$  - a.s. phrases.

Convergence in probability  $\xrightarrow{P}$  and convergence in distribution  $\xrightarrow{D}$  are always relative to the probability measures  $P_{\theta_0}^n$ .

We now introduce the Bayes estimator (under quadratic loss function)

$$(4.5) \quad \hat{\theta}^n = \frac{\int_{\Theta} \theta L^n(\theta, T^n) p(\theta) d\theta}{\int_{\Theta} L^n(\theta, T^n) p(\theta) d\theta}$$

of  $\theta_0$  ; here

$$(4.6) \quad L^n(\theta, T^n) = L_0^n(\theta) \exp \left\{ \int_0^{T^n} \ln[g(\theta)] dN^n(s) + \int_0^{T^n} (1-g(\theta)) Y^n(s) ds \right\}.$$

We have the following asymptotic result for the estimator  $\hat{\theta}^n$ .

Theorem 4.1

$\sqrt{n}(\hat{\theta}^n - \theta_0)$  is asymptotically normal with mean 0 and variance  $\beta^{-1}$ , where

$$\beta = \{[g'(\theta_0)]^2 / g(\theta_0)\} b = \begin{cases} \theta_0^{-1} \int_0^{\infty} y(s) ds & \text{if } g(\theta) = \theta \\ \theta_0^{-3} \int_0^{\infty} y(s) ds & \text{if } g(\theta) = \theta^{-1} \end{cases}.$$

The proof of this theorem is much inspired by Yu.V. Linkov's asymptotic analysis of the Bayes estimator for a diffusion process, see Basawa and Rao (1980) pp. 241-247. Some places we copy his arguments.

The proof is based on some lemmas. Before we state and prove these lemmas we shall introduce some new processes.

Let  $g(\theta)$  and  $p(\theta)$  be defined as 0 for  $\theta \in R - \theta$  and let  $L^n(\theta, T^n)$  be given by (4.6) for all  $\theta \in R$ . Let the process  $\{\gamma^n(t), t \in R\}$  be defined by

$$(4.7) \quad \gamma^n(t) = L^n(\theta_0 + \frac{t}{\sqrt{n}}, T^n) / L^n(\theta_0, T^n).$$

From (4.5) and (4.7) it is easy to see that

$$(4.8) \quad \sqrt{n}(\hat{\theta}^n - \theta_0) = \frac{\int_{-\infty}^{\infty} t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}{\int_{-\infty}^{\infty} \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}.$$

Let

$$M^n(t) = N^n(t) - \int_0^t g(\theta_0) Y^n(s) ds, \quad t \in R.$$

Furthermore, let for  $t \in R$

$$\alpha^n(t) = [\ln g(\theta_0 + \frac{t}{\sqrt{n}}) - \ln g(\theta_0)] M^n(T^n),$$

$$\begin{aligned} \beta^n(t) = & \{ [\ln g(\theta_0 + \frac{t}{\sqrt{n}}) - \ln g(\theta_0)] g(\theta_0) \\ & - [g(\theta_0 + \frac{t}{\sqrt{n}}) - g(\theta_0)] \} \int_0^{T^n} Y^n(s) ds \end{aligned}$$

and

$$\eta^n(t) = \alpha^n(t) + \beta^n(t).$$

Observe that

$$\eta^n(t) = \ln \gamma^n(t).$$

By using Taylor's formula we may write for  $\theta_0 + \frac{t}{\sqrt{n}} \in \Theta$

$$\alpha^n(t) = t v^n(t) a^n,$$

where

$$a^n = \frac{1}{\sqrt{n}} M^n(T^n)$$

$$v^n(t) = g'(\theta_0 + \delta_1 \frac{t}{\sqrt{n}}) / g(\theta_0 + \delta_1 \frac{t}{\sqrt{n}})$$

and  $\delta_1$  is a number (depending on  $t/\sqrt{n}$ ) between 0 and 1.

Furthermore,

$$\beta^n(t) = - \frac{t^2}{2} w^n(t) b^n,$$

where

$$b^n = \frac{1}{n} \int_0^{T^n} Y^n(s) ds,$$

$$w^n(t) = g(\theta_0) [g'(\theta_0 + \delta_2 \frac{t}{\sqrt{n}})]^2 / [g(\theta_0 + \delta_3 \frac{t}{\sqrt{n}}) - g(\theta_0)]^2$$

and  $\delta_2$  and  $\delta_3$  are numbers (depending on  $t/\sqrt{n}$ ) between 0 and 1.

Lemma 4.2. The finite-dimensional distributions of the process  $\{\gamma^n(t)\}$  converge weakly to the finite-dimensional distributions of the process  $\{\gamma(t)\}$  as  $n \rightarrow \infty$ , where

$$\gamma(t) = \exp\{\xi\beta^{\frac{1}{2}}t - \frac{1}{2}\beta t^2\}, \quad -\infty < t < \infty,$$

and  $\xi$  is a standard normal variable (defined on some probability space  $(\chi, A, P)$ ).

Proof. By means of a version of Helland's (1982) martingale central limit theorem 5.3 we shall first establish that

$$(4.9) \quad a^n = \frac{1}{\sqrt{n}} M^n(T^n) \xrightarrow{D} \xi (g(\theta_0)b)^{\frac{1}{2}},$$

where  $\xi$  is a standard normal variable. For each  $n$  there exists

$F_t^n$ -stopping times  $S_i^n$  such that  $S_i^n \uparrow \infty$  as  $i \rightarrow \infty$  and  $\{\frac{1}{\sqrt{n}} M^n(t \wedge S_i^n \wedge T^n)\}$  is a square integrable  $(P_{\theta_0}^n, F_t^n)$ -martingale with

variance process  $\{\frac{1}{n} \int_0^{t \wedge S_i^n \wedge T^n} g(\theta_0) Y^n(s) ds\}$  on  $[0, \infty]$  (cf. Gill (1980), Section 2). Combining (4.2), (4.3) and Helland's Lemma 5.2 we find that for any sequence  $\{i_n\}$  such that  $i_n \rightarrow \infty$  fast enough,

$$\frac{1}{n} \int_0^{S_{i_n}^n \wedge T^n} g(\theta_0) Y^n(s) ds \xrightarrow{P} g(\theta_0)b;$$

here  $S_{i_n}^n \xrightarrow{P} \infty$ . Now, by applying a version of Helland's Theorem

5.3, where  $M_n$  is a martingale and 1 is replaced by  $\infty$ , we find that

$$\frac{1}{\sqrt{n}} M^n(S_{i_n}^n \wedge T^n) \xrightarrow{D} \xi (g(\theta_0)b)^{\frac{1}{2}}$$

(observe that  $\tilde{\sigma}^\varepsilon[M^n](\infty) = I(1/\sqrt{n} > \varepsilon) \frac{1}{n} \int_0^{S_{i_n}^n \wedge T^n} g(\theta_0) Y^n(s) ds = 0$  for

all large  $n$  for all  $\varepsilon > 0$ ). The statement (4.9) follows if we

can show that

$$(4.10) \quad \frac{1}{\sqrt{n}}[M^n(T^n) - M^n(S_{i_n}^n \wedge T^n)] \xrightarrow{P} 0.$$

Using Lengart's (1977) inequality (see Gill (1980), Theorem 2.4.2) we find that for all  $\varepsilon, \eta > 0$ ,

$$P\left\{\frac{1}{\sqrt{n}}[M^n(T^n) - M^n(S_{i_n}^n \wedge T^n)] > \varepsilon\right\} < \frac{\eta}{\varepsilon} + P\left\{\frac{1}{n} \int_{S_{i_n}^n \wedge T^n}^{T^n} g(\theta_0) Y^n(s) ds > \eta\right\}.$$

From this inequality we obtain (4.10) (remember (4.3)). Thus (4.9) holds. Now, define

$$\phi^n(t) = v^n(t) a^{n-\frac{1}{2}} (w^n(t) b^{n-\beta}) t.$$

Note that

$$t\phi^n(t) = \eta^n(t) + \frac{1}{2}\beta t^2.$$

We shall prove the lemma by showing that the finite-dimensional distributions of the process  $\{\phi^n(t)\}$  converge weakly to the finite-dimensional distributions of the process  $\{\phi(t)\}$ , where  $\phi(t) = \xi\beta^{\frac{1}{2}}$ .

Clearly  $v^n(t) \rightarrow g'(\theta_0)/g(\theta_0)$ ,  $w^n(t) \rightarrow [g'(\theta_0)]^2/g(\theta_0)$  and

$$w^n(t) b^n \xrightarrow{P} \beta \quad \text{as } n \rightarrow \infty. \quad \text{It follows that}$$

$$\phi^n(t) \xrightarrow{D} \xi\beta^{\frac{1}{2}}.$$

Next we shall show that

$$(\phi^n(t_1), \phi^n(t_2)) \xrightarrow{D} (\xi\beta^{\frac{1}{2}}, \xi\beta^{\frac{1}{2}}).$$

By Cramer-Wold's theorem it suffices to show that

$$\begin{aligned} a_1 \phi^n(t_1) + a_2 \phi^n(t_2) &= a_1 (\phi^n(t_1) - \phi^n(t_2)) + (a_1 + a_2) \phi^n(t_2) \\ &= a_1 [v^n(t_1) - v^n(t_2)] a^{n-\frac{1}{2}} - a_1 \frac{1}{2} [(w^n(t_1) b^{n-\beta}) t_1 - (w^n(t_2) b^{n-\beta}) t_2] \\ &\quad + (a_1 + a_2) \phi^n(t_2) \quad \text{converges weakly to } (a_1 + a_2) \xi\beta^{\frac{1}{2}}, \text{ where} \end{aligned}$$



$a_1$  and  $a_2$  are arbitrary constant. But this obviously holds. By arguing as above we easily show that

$$(\phi^n(t_1), \dots, \phi^n(t_r)) \xrightarrow{D} (\xi\beta^{\frac{1}{2}}, \dots, \xi\beta^{\frac{1}{2}}).$$

The lemma follows.

It is clear from the definitions of the processes  $\{\gamma^n(t)\}$  and  $\{\gamma(t)\}$  that they have continuous sample paths. We shall now prove that for any fixed  $A > 0$  the processes  $\{\gamma^n(t), t \in [-A, A]\}$  converge in distribution to the process  $\{\gamma(t), t \in [-A, A]\}$  on  $C[-A, A]$  (the space of continuous functions on  $[-A, A]$  with the uniform metric). In other words, the probability measures  $\nu_n$  generated by the processes  $\{\gamma^n(t), t \in [-A, A]\}$  converge weakly to the probability measure  $\nu$  generated by the process  $\{\gamma(t), t \in [-A, A]\}$  on  $C[-A, A]$ . Because of Lemma 4.2, it is sufficient to prove the following lemma (cf. Billingsley (1968), Theorem 8.1-8.2).

Lemma 4.3. For every  $\varepsilon > 0$ ,

$$\lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} P\left\{ \sup_{\substack{|t_1 - t_2| \leq h \\ t_1, t_2 \in [-A, A]}} |\gamma^n(t_2) - \gamma^n(t_1)| > \varepsilon \right\} = 0.$$

Proof. In the following discussion we shall suppose that  $t$  varies over  $[-A, A]$ . Clearly

$$\sup_{|t_1 - t_2| \leq h} |\gamma^n(t_2) - \gamma^n(t_1)| \leq \sup_{|t_2 - t_1| \leq h} |\eta^n(t_2) - \eta^n(t_1)| \exp\left\{ \sup_{|t| \leq A} \eta^n(t) \right\}.$$

Hence for any  $N > 0$ ,

$$\begin{aligned} P\left\{ \sup_{|t_1 - t_2| \leq h} |\gamma^n(t_2) - \gamma^n(t_1)| > \varepsilon \right\} &\leq P\left\{ \sup_{|t| \leq A} \eta^n(t) > N \right\} \\ &+ P\left\{ \sup_{|t_1 - t_2| \leq h} |\eta^n(t_2) - \eta^n(t_1)| > \varepsilon e^{-N} \right\}. \end{aligned}$$

The lemma follows if we can prove that

$$(4.11) \quad \lim_{N \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left\{ \sup_{|t| \leq A} \eta^n(t) > N \right\} = 0$$

and that for every  $\varepsilon > 0$

$$(4.12) \quad \lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} P\left\{ \sup_{|t_1 - t_2| \leq h} |\eta^n(t_2) - \eta^n(t_1)| > \varepsilon \right\} = 0.$$

For all large  $n$  we have

$$\sup_{|t| \leq A} \eta^n(t) \leq AV |a^n| + \frac{1}{2} A^2 W b^n,$$

where

$$\sup_{|t| \leq A} |v^n(t)| \leq V < \infty \quad \text{and} \quad \sup_{|t| \leq A} w^n(t) \leq W < \infty.$$

Thus

$$P\left\{ \sup_{|t| \leq A} \eta^n(t) > N \right\} \leq P\left\{ |a^n| > N/2AV \right\} + P\left\{ b^n > N/A^2 W \right\}.$$

Since  $a^n \xrightarrow{D} \xi(g(\theta_0)b)^{\frac{1}{2}}$  and  $b^n \xrightarrow{P} b$ , it follows that

$$\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} P\left\{ |a^n| > N/2AV \right\} + \lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} P\left\{ b^n > N/A^2 W \right\} = 0.$$

Hence (4.11) holds. Suppose  $|t_1 - t_2| \leq h$ . Then for all large  $n$

$$\begin{aligned} |\eta^n(t_1) - \eta^n(t_2)| &\leq [ |t_1 - t_2| |v^n(t_1)| + |v^n(t_1) - v^n(t_2)| |t_2| ] |a^n| \\ &\quad + \frac{1}{2} [ |t_1^2 - t_2^2| w^n(t_1) + |w^n(t_1) - w^n(t_2)| |t_2^2| ] b^n \\ &\leq [ hV + v_h^n A ] |a^n| + \frac{1}{2} [ 2AhW + w_h^n A^2 ] b^n, \end{aligned}$$

where

$$v_h^n = \sup_{|t'_1 - t'_2| \leq h} |v^n(t'_1) - v^n(t'_2)|$$

and

$$w_h^n = \sup_{|t'_1 - t'_2| \leq h} |w^n(t'_1) - w^n(t'_2)|.$$

It is easy to see that  $v_h^n \rightarrow 0$  and  $w_h^n \rightarrow 0$  as  $n \rightarrow \infty$ . It follows that (for all large  $n$ )

$$P\left\{\sup_{|t_1-t_2|<h} |\eta^n(t_1)-\eta^n(t_2)| > \varepsilon\right\} \leq P\{|a^n| > \varepsilon/3hV\} \\ + P\{v_h^n A |a^n| > \varepsilon/3\} + P\{[AhW + \frac{1}{2}w_h^n A^2] b^n > \varepsilon/3\}.$$

Now it is not difficult to see that (4.12) holds. This completes the proof of the lemma.

Lemma 4.4. For every  $\varepsilon > 0$

$$\lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P\left\{\int_{|t|>A} |t| \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon\right\} = 0.$$

In order to prove this lemma we need the following results (cf. Borgan (1983)): Let

$$\hat{\theta}^n = g^{-1}(N^n(T^n) / \int_0^{T^n} Y^n(s) ds).$$

Then  $\hat{\theta}^n$  is a maximum likelihood estimator, i.e.  $\hat{\theta}^n$  maximizes  $L^n(\theta, T^n)$  or equivalently  $v^n(\theta) \stackrel{\text{def}}{=} \ln L^n(\theta, T^n)$ ;  $v^n(\theta)$  is non-decreasing for  $\theta < \hat{\theta}^n$  and non-increasing for  $\theta > \hat{\theta}^n$ ;

$$(4.13) \quad \hat{\theta}^n \xrightarrow{P} \theta_0$$

(in fact we have  $\sqrt{n}(\hat{\theta}^n - \theta_0) \xrightarrow{D} \xi \beta^{-1/2}$ ) and

$$(4.14) \quad \frac{1}{n}[v^n(\theta) - v^n(\theta_0)] \xrightarrow{P} [(\ln g(\theta) - \ln g(\theta_0))g(\theta_0) - (g(\theta) - g(\theta_0))] b \stackrel{\text{def}}{=} A(\theta)b.$$

For the sake of completeness we shall prove these more or less known results. By differentiating

$v^n(\theta) = C + (\ln g(\theta))N^n(T^n) - g(\theta) \int_0^{T^n} Y^n(s) ds$  ( $C$  is quantity which does not depend on  $\theta$ ) we find that the two first assertions must hold.

Writing

$$\frac{N(T^n)}{\int_0^{T^n} Y^n(s) ds} = \frac{\frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} M^n(T^n)}{\frac{1}{n} \int_0^{T^n} Y^n(s) ds} + g(\theta_0),$$

and then using (4.2) and (4.9) we obtain

$$N^n(T^n) / \int_0^{T^n} Y^n(s) ds \xrightarrow{P} g(\theta_0).$$

The statement (4.13) follows. Writing

$$\frac{1}{n} [V^n(\theta) - V^n(\theta_0)] = (\ln g(\theta) - \ln g(\theta_0)) \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} M^n(T^n) + A(\theta) \frac{1}{n} \int_0^{T^n} Y^n(s) ds,$$

we see that (4.14) holds.

Proof of Lemma 4.4. We shall prove the lemma by proving that for every  $\varepsilon > 0$

$$(4.15) \quad \lim_{n \rightarrow \infty} \sup P \left\{ \int_{\substack{|t| > A \\ |t/\sqrt{n}| > \delta}} |t| \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} = 0$$

and

$$(4.16) \quad \lim_{A \rightarrow \infty} \lim_{n \rightarrow \infty} \sup P \left\{ \int_{\substack{|t| > A \\ |t/\sqrt{n}| < \delta}} |t| \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} = 0,$$

here  $\delta$  is a positive constant such that  $\theta_0 \pm \delta \in \theta$  and

$$\sup_{|\theta - \theta_0| < \delta} p(\theta) < \infty.$$

Note that

$$\gamma^n(t) = \exp\{V^n(\theta_0 + \frac{t}{\sqrt{n}}) - V^n(\theta_0)\}.$$

Now, since  $V^n(\theta)$  is non-decreasing for  $0 < \theta < \hat{\theta}^n$  and non-increasing for  $\theta > \hat{\theta}^n$ , it follows that if  $|\hat{\theta}^n - \theta_0| < \delta$ ,

$$\sup_{|\theta - \theta_0| > \delta} (V^n(\theta) - V^n(\theta_0)) < \max[V^n(\theta_0 + \delta) - V^n(\theta_0), V^n(\theta_0 - \delta) - V^n(\theta_0)].$$

Using this we find that

$$(4.17) \quad P\left\{ \int_{\substack{|t| > A \\ |t/\sqrt{n}| > \delta}} |t| \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} < P\{|\hat{\theta}^n - \theta_0| > \delta\} \\ + P\left\{ \exp\{V^n(\theta_0 + \delta) - V^n(\theta_0)\} \int_{-\infty}^{\infty} |t| p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} \\ + P\left\{ \exp\{V^n(\theta_0 - \delta) - V^n(\theta_0)\} \int_{-\infty}^{\infty} |t| p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\}.$$

Since  $\hat{\theta}^n \xrightarrow{P} \theta_0$ , the first term on the right-hand side of the equality (4.17) converges to zero as  $n \rightarrow \infty$ . Let us now consider the second term. First note that

$$\int_{-\infty}^{\infty} |t| p(\theta_0 + \frac{t}{\sqrt{n}}) dt = n \int_{-\infty}^{\infty} |\theta - \theta_0| p(\theta) d\theta < n \left[ \int_{-\infty}^{\infty} \theta p(\theta) d\theta + \theta_0 \right] \stackrel{\text{def}}{=} nK.$$

Let  $-c = \frac{1}{2}A(\theta_0 + \delta)b$ . Then  $c > 0$  (it is easy to see that  $A(\theta) < 0$  for  $\theta \neq \theta_0$ ) and by (4.14)

$$\frac{1}{n}[V^n(\theta_0 + \delta) - V^n(\theta_0)] \xrightarrow{P} -2c.$$

It follows that

$$P\left\{ \exp\{V^n(\theta_0 + \delta) - V^n(\theta_0)\} \int_{-\infty}^{\infty} |t| p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} \\ < P\{Kn \exp\{-nc\} > \varepsilon\} + P\left\{ \frac{1}{n}[V^n(\theta_0 + \delta) - V^n(\theta_0)] > -c \right\}$$

$\rightarrow 0$  as  $n \rightarrow \infty$ . By copying the above arguments with  $-\delta$  in place of  $+\delta$  we find that also the third term on the right-hand side of the equality (4.17) converges to zero as  $n \rightarrow \infty$ . Thus (4.15) holds.

Let us so prove (4.16). Let

$$v^* = \sup_{|\theta - \theta_0| < \delta} \left| \frac{\ln g(\theta) - \ln g(\theta_0)}{\theta - \theta_0} \right| \quad (= \sup_{|t/\sqrt{n}| < \delta} |v^n(t)|)$$

and

$$w^* = \inf_{|\theta - \theta_0| < \delta} \left\{ \frac{A(\theta)}{-\frac{1}{2}(\theta - \theta_0)^2} \right\} \quad (= \inf_{|t/\sqrt{n}| < \delta} w^n(t)),$$

here  $(\ln g(\theta) - \ln g(\theta_0))/(\theta - \theta_0)$  and  $A(\theta)/(-\frac{1}{2})(\theta - \theta_0)^2$  are defined as  $g'(\theta_0)/g(\theta_0)$  and  $(g'(\theta_0))^2/g(\theta_0)$  respectively for  $\theta = \theta_0$ . Note that  $v^* < \infty$  and  $0 < w^*$ . It follows that for  $|t|/\sqrt{n} \leq \delta$ ,

$$\eta^n(t) = \ln \gamma^n(t) = t v^n(t) a^{n-\frac{t^2}{2} w^n(t) b^n} \leq |t| v^* |a^n|^{-\frac{t^2}{2} w^* b^n}.$$

Since

$$b - |b^n - b| \leq b^n$$

we must have

$$\eta^n(t) \leq |t| v^* |a^n|^{-\frac{1}{2}(b - |b^n - b|) w^* t^2}$$

for  $|t|/\sqrt{n} \leq \delta$ . It follows that for any  $N > 0$  and  $0 < b_0 < b$ ,

$$\begin{aligned} P\left\{ \int_{\substack{|t| > A \\ |t|/\sqrt{n} \leq \delta}} |t| \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon \right\} &\leq P\{|a^n| > N\} + P\{|b^n - b| > b_0\} \\ &+ P\left\{ \sup_{|\theta - \theta_0| \leq \delta} p(\theta) \int_{|t| > A} |t| \exp\{|t| v^* N - \frac{1}{2}(b - b_0) w^* t^2\} dt > \varepsilon \right\}. \end{aligned}$$

Now by using that  $a^n \xrightarrow{D} \xi(g(\theta_0)b)^{\frac{1}{2}}$ ,  $b^n \xrightarrow{P} b$  and the fact that

$$\lim_{A \rightarrow \infty} \int_{|t| > A} |t| \exp\{|t| v^* N - \frac{1}{2}(b - b_0) w^* t^2\} dt = 0,$$

we easily obtain (4.16). This completes the proof of Lemma 4.4.

Proof of Theorem 4.1. Relation (4.8) says that

$$(4.18) \quad \sqrt{n}(\hat{\theta}^n - \theta_0) = \frac{\int_{-\infty}^{\infty} t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}{\int_{-\infty}^{\infty} \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds}$$

Since the finite-dimensional distributions of the process  $\{\gamma^n(t)\}$  converge weakly to the finite-dimensional distributions of the process  $\{\gamma(t)\}$ , where  $\gamma(t) = \exp\{\xi \beta^{\frac{1}{2}} t - \frac{1}{2} \beta t^2\}$  ( $\xi$  is a standard normal variable), we obtain by passing to the limit formally in (4.18),

$$\sqrt{n}(\hat{\theta}^n - \theta_0) \xrightarrow{D} \frac{\int_{-\infty}^{\infty} t \gamma(t) dt}{\int_{-\infty}^{\infty} \gamma(s) ds} = \frac{\sqrt{2\pi} \beta^{-1} \xi \exp\{\frac{1}{2} \xi^2\}}{\sqrt{2\pi} \beta^{-\frac{1}{2}} \exp\{\frac{1}{2} \xi^2\}} = \beta^{-\frac{1}{2}} \xi.$$

We shall now justify this limiting argument. Let  $0 < A < \infty$ . For any fixed  $k_1$  and  $k_2$  the functional

$$f_{k_1, k_2}(x) = k_1 \int_{-A}^A s x(s) ds + k_2 \int_{-A}^A x(s) ds$$

defined on  $C[-A, A]$  is continuous. Lemma (4.2) and (4.3) imply that the processes  $\{\gamma^n(t), t \in [-A, A]\}$  converge in distribution to the process  $\{\gamma(t), t \in [-A, A]\}$ . Hence the distribution of

$f_{k_1, k_2}(\gamma^n)$  converges weakly to the distribution of  $f_{k_1, k_2}(\gamma)$  as  $n \rightarrow \infty$  (cf. Billingsley (1968) p. 30). From this fact and the continuity of  $p(\theta)$  at  $\theta_0$ , it follows that

$$(4.19) \quad \begin{pmatrix} \int_{-A}^A t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt \\ \int_{-A}^A \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \int_{-A}^A t \gamma(t) p(\theta_0) dt \\ \int_{-A}^A \gamma(t) p(\theta_0) dt \end{pmatrix}$$

Define

$$X_A^n = \frac{\int_{-A}^A t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}{\int_{-A}^A \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds}.$$

Then by (4.19)

$$(4.20) \quad X_A^n \xrightarrow{D} X_A \stackrel{\text{def}}{=} \frac{\int_{-A}^A t \gamma(t) dt}{\int_{-A}^A \gamma(s) ds}.$$

Furthermore,

$$(4.21) \quad X_A \rightarrow \frac{\int_{-\infty}^{\infty} t \gamma(t) dt}{\int_{-\infty}^{\infty} \gamma(s) ds} = \beta^{-\frac{1}{2}} \xi \quad \text{as } A \rightarrow \infty.$$

The conclusion of the theorem now follows by an application of Theorem 25.5 of Billingsley (1979) if we can show that for every  $\varepsilon > 0$

$$(4.22) \quad \lim_{A \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{|Y_A^n| > \varepsilon\} = 0,$$

where

$$Y_A^n = \sqrt{n}(\hat{\theta}^n - \theta_0) - X_A^n.$$

In order to show (4.22) we write  $Y_A^n$  in the following way:

$$Y_A^n = X_A^n \left\{ \frac{1 + \frac{\int_{|t|>A} t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}{\int_{-A}^A s \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds}}{1 + \frac{\int_{|t|>A} \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt}{\int_{-A}^A \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds}} - 1 \right\}.$$

Now, clearly

$$\begin{aligned} P\{|Y_A^n| > \varepsilon\} &\leq P\{|X_A^n| > N\} + P\left\{\left|\int_{-A}^A s \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds\right| < \delta_1\right\} \\ &+ P\left\{\int_{-A}^A \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds < \delta_2\right\} \\ &+ P\{|Y_A^n| > \varepsilon, |X_A^n| \leq N, \left|\int_{-A}^A s \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds\right| > \delta_1, \int_{-A}^A \gamma^n(s) p(\theta_0 + \frac{s}{\sqrt{n}}) ds > \delta_2\} \end{aligned}$$

for all  $\delta_1 > 0$ ,  $\delta_2 > 0$  and  $N > 0$ . Furthermore, for fixed  $\delta_1, \delta_2$  and  $N$  there exists positive numbers  $\varepsilon_1$  and  $\varepsilon_2$  (not depending on  $n$  and  $A$ ) such that

$$\begin{aligned} P\{|Y_A^n| > \varepsilon, |X_A^n| \leq N, \dots\} &\leq P\left\{\left|\int_{|t|>A} t \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt\right| > \varepsilon_1\right\} \\ &+ P\left\{\int_{|t|>A} \gamma^n(t) p(\theta_0 + \frac{t}{\sqrt{n}}) dt > \varepsilon_2\right\} \end{aligned}$$

(choose  $\varepsilon_1$  and  $\varepsilon_2$  such that  $\varepsilon_1 < \delta_1 \varepsilon / N$  and

$(1 - \varepsilon_1 / \delta_1) / (1 + \varepsilon_2 / \delta_2) > 1 - \varepsilon / N$ ). Using the above inequalities, (4.19)-

(4.21) and Lemma 4.4, we find that



$$\begin{aligned} & \lim_{A \rightarrow \infty} \sup \lim_{n \rightarrow \infty} P\{|Y_A^n| > \varepsilon\} \leq P\{\beta^{-1/2} |\xi| > N\} \\ & + P\left\{\left|\int_{-\infty}^{\infty} s \gamma(s) p(\theta_0) ds\right| = \sqrt{2\pi\beta}^{-1} |\xi| \exp\{\frac{1}{2}\xi^2\} p(\theta_0) \leq \delta_1\right\} \\ & + P\left\{\left|\int_{-\infty}^{\infty} \gamma(s) p(\theta_0) ds\right| = \sqrt{2\pi\beta}^{-1/2} \exp\{\frac{1}{2}\xi^2\} p(\theta_0) \leq \delta_2\right\}. \end{aligned}$$

By letting  $N \rightarrow \infty$ ,  $\delta_1 \rightarrow 0$  and  $\delta_2 \rightarrow 0$ , (4.22) follows. This completes the proof of the theorem.

Finally in this section we shall give some comments concerning Theorem 4.1 when  $p$  is a natural conjugate prior (see Section 3). Suppose that  $g(\theta) = \theta$  and  $p(\theta) = d^c \theta^{c-1} e^{-d\theta} / \Gamma(c)$ , i.e.  $p$  is a gamma density with parameters  $c$  and  $d$ . Then

$$\hat{\theta}^n = \frac{c + N^n(T^n)}{d + \int_0^{T^n} Y^n(s) ds}$$

and by Theorem 4.1

$$\sqrt{n}(\hat{\theta}^n - \theta_0) \xrightarrow{D} \xi \beta^{-1/2},$$

where  $\xi$  is a standard normal variable. This asymptotically result can in fact be established directly by writing

$$\sqrt{n}(\hat{\theta}^n - \theta_0) = \frac{\frac{c}{\sqrt{n}} + \frac{1}{\sqrt{n}}(N^n(T^n) - \int_0^{T^n} \theta_0 Y^n(s) ds) - \frac{d\theta_0}{\sqrt{n}}}{\frac{d}{n} + \frac{1}{n} \int_0^{T^n} Y^n(s) ds}$$

and then using (4.2) and (4.9).

Similarly, if  $g(\theta) = \theta^{-1}$  and  $p$  is an inverted gamma density with parameters  $c$  and  $d$ , we can establish directly that

$$\sqrt{n}(\hat{\theta}^n - \theta_0) = \sqrt{n}\left(\frac{d + \int_0^{T^n} Y^n(s) ds}{c - 1 + N^n(T^n)} - \theta_0\right) \xrightarrow{D} \xi \beta^{-1/2}.$$

### 5. A special case

In this section we shall motivate the set-up of Section 4 by presenting a special case. In particular we shall draw attention to the assumptions (4.2) and (4.3).

Consider the set-up of Section 4. Assume that

$$N^n(t) = \sum_{i=1}^n I(X_i^n \leq t),$$

$$Y^n(t) = \sum_{i=1}^n I(X_i^n > t) \quad (= n - N^n(t^-)),$$

$$F_t^n = \sigma(\text{all } Q^n\text{-null sets}) \vee \sigma(N^n(u), u \leq t)$$

and

$$T^n = \min(t_0, X_{(r_n)}^n);$$

here the  $X_i^n$ 's are independent and exponentially distributed random variables with failure rate  $g(\theta)$  ( $=\theta$  or  $\theta^{-1}$ ) and expectation  $g(\theta)^{-1}$  ( $=\theta^{-1}$  or  $\theta$ ) under  $P_\theta^n$ ,  $t_0$  and  $r_n$  are constants,  $t_0 \in (0, \infty]$ ,  $r_n \in \{1, 2, \dots, n\}$ , and  $X_{(r_n)}^n$  is the  $r_n$ -th smallest  $X_i^n$ . By using Brémaud and Jacod's (1977) proposition p. 373 we easily verify that  $\{N^n(t)\}$  has  $(P_\theta^n, F_t^n)$ -intensity  $\{g(\theta)Y^n(t)\}$ . We can give the following interpretation of the model in this case:  $n$  items are put on test at time 0. The lifelengths (time to failures) of the items are independent and exponentially distributed random variables with failure rate  $g(\theta)$  for a  $\theta \in \Theta$ . The testing is stopped at  $T^n = \min(t_0, X_{(r_n)}^n)$ . The random variable  $N^n(t)$  represents the number of failures in  $[0, t]$ ,  $Y^n(t)$  represents the number of non-failed items just before time  $t$  and  $\int_0^{T^n} Y^n(t) dt$  represents the total time on test.

Assume now that

$$r_n/n \rightarrow r \quad \text{as } n \rightarrow \infty,$$

where

$$0 < r \leq 1.$$

We shall show that (4.2) and (4.3) hold with

$$y(t) = \bar{F}(t)I(F(t) \leq r)I(t \leq t_0) \quad (= \bar{F}(t)I(t \leq \min(t_0, F^{-1}(r)))),$$

where

$$\bar{F}(t) = 1 - F(t) = e^{-g(\theta_0)t}.$$

We shall establish (4.2) by showing (4.4).

Let

$$F_n(t) = 1 - \bar{F}_n(t) = N^n(t)/n.$$

Now, by using the Glivenko-Cantelli theorem we find that for

$$0 \leq t < \infty$$

$$\int_0^t [Y^n(s)/n] I(s \leq T^n) ds = \int_0^t \bar{F}_n(s) I(F_n(s) \leq r_n/n) I(s \leq t_0) ds \xrightarrow{P} \int_0^t y(s) ds,$$

hence (4.4) holds. The statement (4.3) is seen to hold by noting that

$$\begin{aligned} P\left\{ \int_{t \wedge T^n}^{T^n} [Y^n(s)/n] ds > \varepsilon \right\} &\leq P\left\{ \int_t^\infty [Y^n(s)/n] ds > \varepsilon \right\} = P\left\{ \int_t^\infty \bar{F}_n(s) ds > \varepsilon \right\} \\ &\leq \frac{1}{\varepsilon} E \int_t^\infty \bar{F}_n(s) ds = \frac{1}{\varepsilon} \int_t^\infty E \bar{F}_n(s) ds = \frac{1}{\varepsilon} \int_t^\infty \bar{F}(s) ds \rightarrow 0 \quad \text{as } t \rightarrow \infty. \end{aligned}$$

# Appendix.

We shall prove that  $L(\theta, T)$  is a  $B \times F_T$ -measurable function on  $\Theta \times \Omega$ . It suffices to show that

$$(A.1) \quad \int_0^{T(\omega)} \ln \mu_i(\theta, s, \omega) N_i(ds, \omega) \text{ is } B \times F_T\text{-measurable}$$

and

$$(A.2) \quad \int_0^{T(\omega)} \lambda_i(\theta, s, \omega) ds \text{ is } B \times F_T\text{-measurable.}$$

We shall here prove (A.1) only; the proof of (A.2) is similar.

Now, to prove (A.1) it is sufficient to show that

$$\int_0^{T(\omega)} v(\theta, s, \omega) N_i(ds, \omega) \text{ is } B \times F_T\text{-measurable}$$

whenever  $v$  is a non-negative  $B \times P(F_t)$ -measurable function on  $\Theta \times [0, \infty) \times \Omega$ . Furthermore, since  $P(F_t)$  is generated by the rectangles  $\{0\} \times A$ ,  $A \in F_0$  and  $[u, t] \times A$ ,  $0 < u < t$ ,  $A \in F_u$ , (see e.g. Dellacherie and Meyer (1978) p. 125) and  $B \times P(F_t)$  is generated by  $\{B \times F, B \in B, F \in P(F_t)\}$ , it is sufficient to show that for  $B \in B$  and  $A \in F_u$

$$(A.3) \quad I_B(\theta) \int_0^{T(\omega)} I_A(\omega) I_{[u, t]}(s) N_i(ds, \omega) \text{ is } B \times F_T\text{-measurable.}$$

But (A.3) is easily shown (cf. Meyer (1965), p. 109).

# Acknowledgement.

The author thanks Bent Natvig for helpful comments.

## References

- Aalen, O.O. (1975). Statistical inference for a family of counting processes. Ph. D. diss, Dept. of Statistics, Univ. of California, Berkeley.
- Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* 6, 701-726.
- Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Inst. Statist. Rev.* 50, 319-258.
- Andersen, P.K., and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* 10, 1100-1120.
- Ash, R.B. (1982). *Real Analysis and Probability*. Academic Press, New York.
- Aven, T. (1983). A theorem for determining the compensator of a counting process. Unpublished paper.
- Barlow, R.E. and Proschan, F. (1979). Inference for the exponential life distribution. Research Report. Operat. Res. Center, Univ. of California, Berkeley.
- Barlow, R.E. and Proschan, F. (1980). Life distribution models and incomplete data. Research Report. Operat. Res. Center., Univ. of California, Berkeley.
- Barra, J.-R. (1981). *Mathematical Basis of Statistics*. Academic Press. New York.
- Basawa, I.V., and Prakasa Rao, B.L.S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, New York.
- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Billingsley, P. (1979). *Probability and Measure*. Wiley, New York.
- Borgan, Ø. (1983). Maximum likelihood estimation in a parametric counting process model, with applications to censored failure time data and multiplicative models. Research Report, Dept. of Math. and Statistics, Agricultural Univ. of Norway.
- Brémaud, P., and Jacod, J. (1977). Processus ponctuels et martingales: Résultats récents sur la modélisation et le filtrage. *Adv. Appl. Prob.* 9, 362-416.
- Brémaud, P. (1981). *Point Processes and Queues*. Springer Verlag, New York.
- Dellacherie, C. and Meyer, P.A. (1978). *Probabilities and Potential*. North-Holland, Amsterdam.
- Gill, R.D. (1980). Censoring and Stochastic Integrals. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.
- Helland, I.S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* 9, 79-94.
- Lenglart, E. (1977). Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré* 13, 171-179.
- Meyer, P.A. (1966). *Probability and Potentials*, Blaisdell, Waltham, Mass.
- Rebolledo, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille de processus ponctuels. In *Lecture Notes in Mathematics*, 636, 27-70. Springer Verlag, Berlin.
- Rebolledo, R. (1980). Central limit theorems for local martingales. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 51, 269-286.

